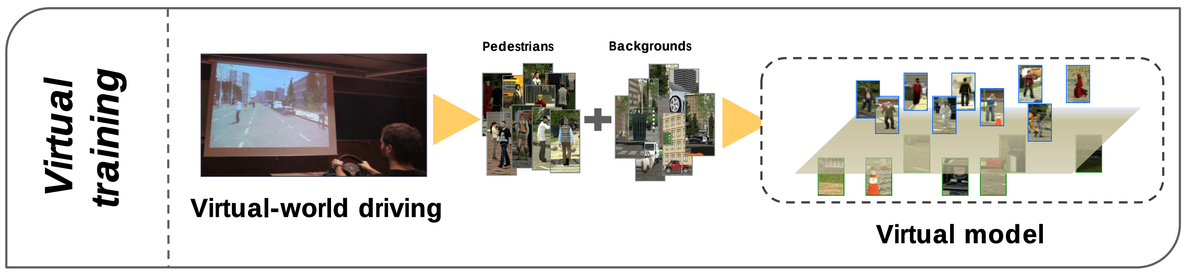
Weakly Supervised Automatic Annotation of Pedestrian Bounding Boxes

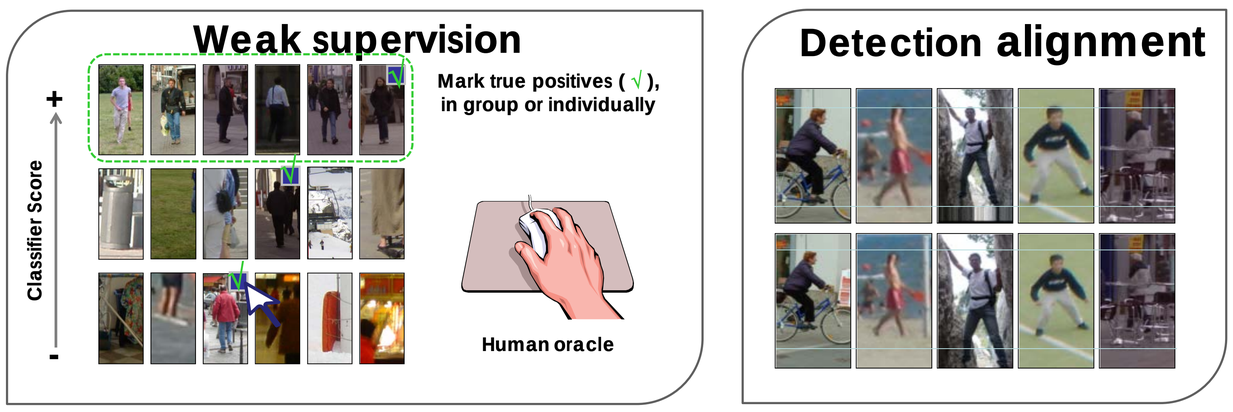
D. Vázquez[1], J. Xu[1], S. Ramos[1], A. M. López[1,2], D. Ponsa[1,2]
dacv@cvc.uab.es
www.cvc.uab.es/domainadaptation

[1]Computer Vision Center [2]Dept. Computer Science
Autonomous University of Barcelona

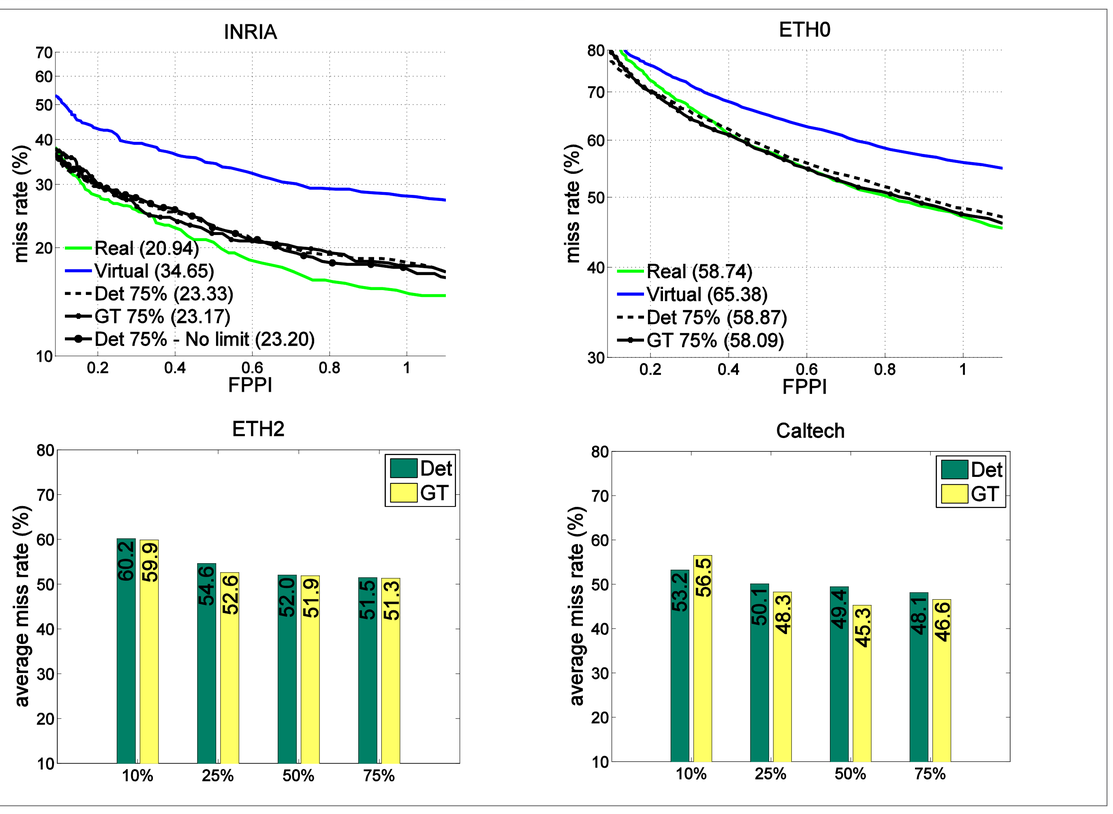
Context: vision-based human detection

Pedestrian detectors rely on classifiers trained with annotated samples. The annotation step is an intensive and subjective task when it has to be done by persons. Therefore, it is worth to minimize the human intervention in such a task by using automatically generated synthetic data. In this work, we propose *a* strategy that consists on collecting real-world samples by using a classifier trained with synthetic data. After, a human oracle rejects the false detections by an efficiently weak annotation. Finally, a new classifier is trained with the accepted detections. We show that this classifier is competitive with respect to the counterpart, which is trained with samples collected by manually annotating hundreds of pedestrian bounding boxes.



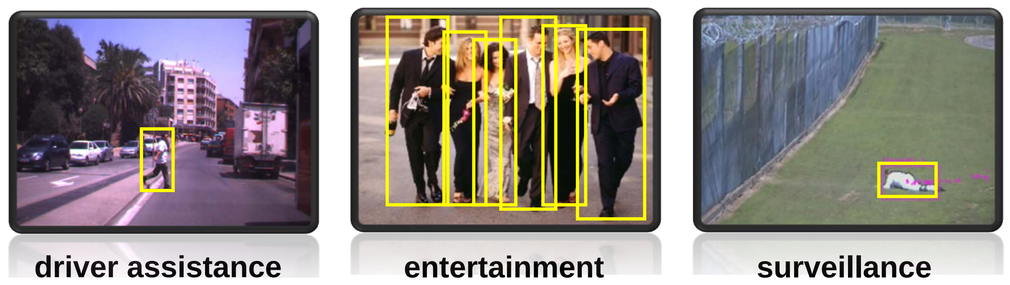


Results



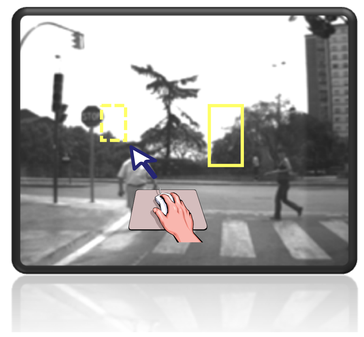
Context: vision-based human detection

Vision-based human detection is of paramount relevance for many applications.



Problem (1): tiresome manual annotations

Developing successful human detectors involves tiresome and subjective manual annotations.

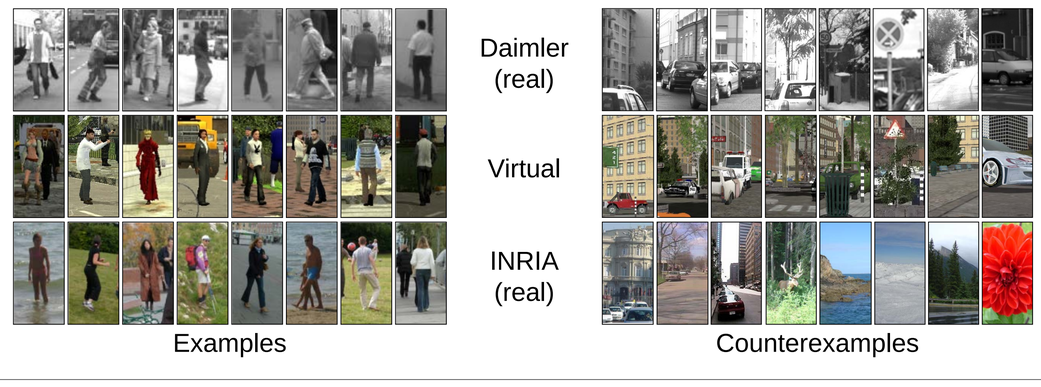


Our proposal [1]: training with virtual-world samples

Train based on *virtual-world* samples for operating in *real-world* images.

Problem (2): dataset shift [2]

1. Virtual and real data come from different *eyes*.
2. Poses, clothes and background change with the environment.



Our proposal: Automatic annotation

- Train with virtual-world data
- Detecting in real-world images and weakly annotating the detections
- Retrain with real-world detections

References

[1] J. Marín, D. Vázquez, D. Gerónimo, and A.M. López. *Learning appearance in virtual scenarios for pedestrian detection*. Conf. on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 2010.
[2] D. Vázquez, A. M. López, J. Marín, D. Ponsa. *Virtual worlds and active learning for human detection*. International Conference on Multimodal Interaction (ICMI), Alicante, Spain, 2011.

Acknowledgements

This work is supported by the Spanish MICINN projects TRA2011-29454-C03-01 and TIN2011-29494-C03-02, the Chinese Scholarship Council (CSC) grant No.2011611023 and Sebastian Ramos' FPI Grant BES-2012-058280.